

VŠB – Technická univerzita Ostrava
Fakulta elektrotechniky a informatiky
Katedra informatiky

Detekce významných oblastí v obrazech

Salient Region Detection in Images

Zadání bakalářské práce

Student: **Jakub Kajfosz**

Studijní program: B2647 Informační a komunikační technologie

Studijní obor: 2612R025 Informatika a výpočetní technika

Téma: Detekce významných oblastí v obrazech
Salient Region Detection in Images

Jazyk vypracování: čeština

Zásady pro vypracování:

Při pozorování reálných obrazů je člověk schopen rozlišit oblasti zájmů od nedůležitého pozadí. Automatické nalezení takovýchto oblastí by mohlo pomoci i v mnoha aplikacích analýzy obrazu, např. k urychlení detekce a rozpoznání objektů v obrazech. V posledních letech vzniklo mnoho algoritmů pro detekci významných oblastí v obrazech používajících různé přístupy. Cílem práce je vybrat jednu nebo více ze známých metod a ověřit jejich úspěšnost na vybrané množině obrazů.

Ve své práci proveďte:

1. Představte přehled existujících metod detekce významných oblastí v obrazech.
2. Vámi vybranou metodu popište a naimplementujte v jazyce C/C++ nebo Python.
3. Otestujte metodu na množině obrazů z veřejného datasetu.
4. Zhodnoťte dosažené výsledky.

Seznam doporučené odborné literatury:

- [1] M. M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, S. M. Hu, "Global Contrast Based Salient Region Detection," in IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 37, no. 3, pp. 569-582, 2015
- [2] J. Zhang, S. Sclaroff, Z. Lin, X. Shen, B. Price, R. Měch. "Minimum Barrier Salient Object Detection at 80 FPS." in Proc. IEEE International Conference on Computer Vision (ICCV), 2015
- [3] X. Hou, L. Zhang. "Saliency detection: A spectral residual approach." In Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, pages 1-8. IEEE, 2007


Formální náležitosti a rozsah bakalářské práce stanoví pokyny pro vypracování zveřejněné na webových stránkách fakulty.

Vedoucí bakalářské práce: **Ing. Michael Holuša, Ph.D.**


Datum zadání: 01.09.2018

Datum odevzdání: 30.04.2020





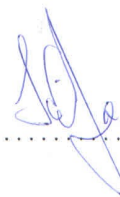
doc. Ing. Jan Platoš, Ph.D.
vedoucí katedry



prof. Ing. Pavel Brandštetter, CSc.
děkan fakulty

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně. Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

V Ostravě 15. května 2020



.....

Tímto bych rád poděkoval vedoucímu bakalářské práce panu Ing. Michaelu Holušovi, Ph.D. za jeho podnětné a konstruktivní rady, které mi výrazně pomohly při tvorbě této práce.

Abstrakt

Tato bakalářská práce řeší detekci významných oblastí v obrazech. Při pozorování prostředí je člověk schopen odlišit oblasti zájmu od nedůležitého pozadí. Automatické nalezení takovýchto oblastí by mohlo pomoci v mnoha aplikacích analýzy obrazu, např. k urychlení detekce a rozpoznání objektů v obrazech. Cílem této práce je analyzovat a porovnat chování tří známých metod pro detekci významných oblastí a ověřit jejich úspěšnost na vybrané množině obrazů. V první části této práce představíme přehled existujících metod detekce významných oblastí v obrazech, následně popíšeme a analyzujeme tři z těchto metod. Na závěr tyto metody otestujeme na množině obrazů z veřejného datasetu MSRA10K a porovnáme výsledky.

Klíčová slova: významné oblasti; obraz; detekce

Abstract

This bachelor's thesis studies the detection of salient regions in images. When observing the environment, one is able to distinguish areas of interest from an unimportant background. Automatic detection of such areas could help in many image analysis applications, such as speeding up the detection and recognition of objects in images. The aim of this work is to analyze and compare the behavior of three known methods for the detection of salient regions and to verify their success on a selected set of images. In the first part of this work we present an overview of existing methods for salient region detection in images, then we describe and analyze three of these methods. Finally, we test these methods on a set of images from the public MSRA10K dataset and compare the results.

Keywords: salient regions; image; detection

Obsah

Seznam použitých zkratek a symbolů	8
Seznam obrázků	9
Seznam tabulek	10
1 Úvod	11
1.1 Cíle práce	11
2 Přehled existujících metod detekce významných oblastí v obrazech	12
2.1 Predikce vizuální fixace	12
2.2 Detekce významných oblastí	12
3 Výběr metod	14
4 Metoda histogramového kontrastu	15
4.1 Zrychlení založené na histogramu	15
4.2 Vyhlazení barevného prostoru	16
4.3 Implementační detaily	17
5 Metoda minimální bariérové vzdálenosti	18
5.1 Transformace vzdálenosti	18
5.2 Rychlá transformace MBD pomocí rastrového skenování	19
5.3 Transformace MBD pro detekci významných objektů	21
5.4 Post-processing	21
6 Metoda spektrálního rezidua	22
6.1 Od spektrálního rezidua k mapě významnosti	22
6.2 Výběr vizuální škály	24
6.3 Reakce na psychologické vzorce	24
7 Porovnání	25
7.1 Precision-recall křivka	25
7.2 Rychlost	26
7.3 Různé kategorie obrazů	26
8 Závěr	35
Literatura	36

Seznam použitých zkratek a symbolů

HC	– Histogram-based Contrast
MBD	– Minimum Barrier Distance
SR	– Spectral Residual
RGB	– Red Green Blue
CBIR	– Content-based Image Retrieval
FT	– Frequency-tuned
CA	– Context-aware
RC	– Region-based Contrast
GC	– Global Cues
SF	– Saliency Filters
HDCT	– High-Dimensional Color Transform
BMS	– Boolean Map-based Saliency
RBD	– Robust Background Detection
GMR	– Graph-based manifold ranking
HS	– Hierarchical Saliency
LR	– Low rank
SMD	– Structured matrix decomposition
ST	– Saliency Tree
PCA	– Principal Component Analysis
CNN	– Convolutional Neural Network
MSRA	– Microsoft Research Asia
FPS	– Frames per second
OpenCV	– Open Computer Vision Library
ROC	– Receiver operating characteristic

Seznam obrázků

1	Pro vstupní obrázek (vlevo) vypočítáme jeho barevný histogram (uprostřed). Odpovídající barvy zásobníků histogramu jsou zobrazeny ve spodní liště. Kvantifikovaný obrázek (vpravo) používá pouze 43 barev histogramu a stále si zachovává dostatečnou vizuální kvalitu pro detekci významnosti.	16
2	Významnost každé barvy v histogramu před (vlevo) a po (vpravo) vyhlazení barevného prostoru, spolu s odpovídajícími mapami významnosti.	17
3	Ilustrace rastrového skenovacího průchodu a inverzního rastrového skenovacího průchodu. Zelený pixel je aktuálně navštívený pixel a jeho maskovaná sousední oblast je znázorněna červeně. [35]	19
4	Mapa významnosti před (uprostřed) a po (vpravo) post-processing operacích . .	21
5	Ilustrace efektu velikosti místního průměrného filtru.	23
6	Ilustrace efektu vizuálního měřítka. Původní obraz (vlevo) je transformován na mapu významnosti po škálování na 32x32 (uprostřed) a 256x256 (vpravo). Můžeme vidět, že ve větším měřítku je jako objekt vnímám celý květ, zatímco v menším měřítku je objektem střed květu.	24
7	Reakce na psychologické vzorce [44]	24
8	Precision-recall křivky	25
9	Porovnání metod na obrazech květin	27
10	Porovnání metod na obrazech zvířat	28
11	Porovnání metod na obrazech koček	29
12	Porovnání metod na obrazech všedních objektů	30
13	Porovnání metod na obrazech lidí	31
14	Porovnání metod na obrazech aut	32
15	Porovnání metod na obrazech dopravních značek	33
16	Porovnání metod na obrazech objektů dotýkajících se okrajů	34

Seznam tabulek

1	Porovnání rychlostí jednotlivých metod - experimentální výsledky	26
---	--	----

1 Úvod

Lidé často a bez námahy posuzují důležitost oblastí ve svém okolí a zaměřují svou pozornost pouze na důležité části. Automatické detekování těchto významných oblastí počítačovými systémy zůstává důležitým cílem, jelikož umožňuje preferenční alokaci výpočetních zdrojů v následné obrazové analýze. Extrahované mapy významnosti jsou široce používány v mnoha aplikacích počítačového vidění, včetně segmentace obrazu na části s objekty zájmu [1, 2], rozpoznávání objektů [3], adaptivní komprese obrazů [4] a editace obrazu se zohledněním obsahu [5, 6, 7, 8].

Významnost vzniká z vizuální jedinečnosti, nepředvídatelnosti, vzácnosti nebo překvapení a je často připisována variacím v atributech obrazů, jako je barva, gradient, hrany a hranice. Vizuální význam, úzce související s tím, jak vnímáme a zpracováváme vizuální podněty, zkoumají různé obory včetně kognitivní psychologie [9, 10], neurobiologie [11, 12] a počítačového vidění [13, 14]. Teorie lidské pozornosti předpokládají, že lidský vizuální systém zpracovává pouze některé části okolí podrobně, zatímco ostatní části zůstávají téměř nezpracovány.

Cílem detekce významných objektů je vypočítání a nalezení mapy významnosti, která zvýrazňuje hlavní objekty a potlačuje pozadí scény. V poslední době tento problém získal díky své užitečnosti velký zájem o výzkum v mnoha aplikacích počítačového vidění, např. detekce objektů, rozpoznávání dění a různé aplikace pro zpracování obrazu / videa. Vzhledem k rozvíjejícím se aplikacím na mobilních zařízeních a rozsáhlým datasetům by metody detekce významných objektů neměly pouze vytvářet vysoce kvalitní mapy významnosti, ale měly by být také vysoce výpočetně efektivní.

1.1 Cíle práce

Cílem této práce je analyzovat a porovnat chování tří známých metod pro detekci významných oblastí a ověřit jejich úspěšnost na vybrané množině obrazů.

V první části této práce představíme přehled existujících metod detekce významných oblastí v obrazech, následně popíšeme a analyzujeme tři z těchto metod. První metoda se zaměřuje na využití histogramového kontrastu (HC), zatímco druhá využívá transformaci přibližné minimální bariérové vzdálenosti (MBD), a třetí aplikuje logaritmické spektrum pro získání spektrálního rezidua (SR). Na závěr tyto metody otestujeme na množině obrazů z veřejného datasetu MSRA10K [15] a porovnáme výsledky.

2 Přehled existujících metod detekce významných oblastí v obrazech

V minulých letech bylo představeno mnoho metod pro detekci významných objektů a oblastí. Existuje celá řada technik vizuální salience, z nichž většina spadá do dvou hlavních kategorií: predikce vizuální fixace a detekce významných oblastí.

2.1 Predikce vizuální fixace

V minulosti byla predikce fixace důležitou větví detekce významnosti. Tato technika využívá kontrast prvků nízké úrovně, jako je barva, orientace a intenzita. V roce 1998 Itti a kol. [13] představili metodu kontrastu zaměřeného na střed pro detekci významnosti pomocí barev, intenzity a orientačních rysů. Ma a Zhang [16] navrhli novou metodu pro generování map významnosti založenou na analýze lokálního kontrastu. Na základě současného porozumění chování lidského vizuálního systému, Meur a kol. [17] navrhli ucelený výpočetní systém k modelování vizuální pozornosti zdola nahoru. Borji a Itti [18] představili model salience, který zvažuje lokální a globální rarity obrazových patchů jako dva komplementární procesy. Harel a kol. [19] navrhli grafové schéma vizuální významnosti pro detekci významných oblastí. Frintrop a kol. [20] upravili model výběhu Itti a kol. [13] zavedením dvojité pyramidy pro výpočet gaussovských rozdílů a využili ji pro segmentaci obrazu na části s objekty zájmu.

2.2 Detekce významných oblastí

Metody detekce významných objektů založené na globálních attributech prokázaly výhody v některých úlohách. Lze je obecně rozdělit do dvou kategorií: metody využívající model zdola nahoru a model shora dolů.

2.2.1 Modely zdola nahoru:

Tyto modely používají pouze nízkoúrovňové vlastnosti pro detekci významnosti. Achanta a kol. [21] představili metodu segmentace pro detekci významných oblastí. Liu a kol. [22] vypočítali lokální kontrast hran a objem barev v barevném prostoru HSV pro detekci významných oblastí a použili další zpracování ve srovnání s tradičním vyhledáváním obrazů podle obsahu (CBIR). Achanta a kol. [14] zavedli metodu frekvenčního vyladění (FT) pro detekci významných oblastí. Tato metoda využívá barevné a jasové vlastnosti a může vytvářet mapy významnosti v plném rozlišení s dobře definovanými hranicemi pro významné objekty. Goferman a kol. [23] představili kontextovou (CA) metodu detekce významnosti, která si klade za cíl detekovat obrazové oblasti představující scénu. Cheng a kol. [24, 25, 26] navrhli metodu histogramového kontrastu (HC) a využili globální podněty (GC) a oblastní kontrast (RC) pro detekci významnosti. Mezitím, Perazzi a kol. [27] navrhli použití salienčních filtrů (SF) pro detekci významných oblastí. Kim

a kol. [28] představili novou techniku, která automaticky detekuje hlavní oblasti obrazu pomocí barevné transformace ve vysokém rozlišení (HDCT). Zhang a Sclaroff [29] navrhli nový booleanový mapový model (BMS). Li a kol. [30] prozkoumali využití světelných polí pro detekci významnosti.

2.2.2 Modely shora dolů:

Modely shora dolů využívají vysokoúrovňové vlastnosti nebo algoritmy strojového učení specificky zaměřené na detekci významnosti.

a) Modely využívající vysokoúrovňové vlastnosti: Wei a kol. [31] navrhli geodetickou metodu, která využívá vlastnosti pozadí. Jiang a kol. [32] zvažovali odchylku a prostorové rozložení významných objektů a pozadí. Zhu a kol. [33] představili novou metodu měření pozadí, která používá intuitivní geometrickou interpretaci pro detekci pozadí s vysokou přesností (RBD). Jiang a kol. [34] navrhli nový segmentační algoritmus k integraci významných nízkoúrovňových podnětů a objektových vlastností tvarů. Zhang a kol. [35] představili algoritmus pro transformaci přibližné minimální bariérové vzdálenosti (MBD) pro detekci významných oblastí. Yang a kol. [36, 37] navrhli metodu varietního pořadí založeného na grafech (GMR) pro extrakci pozadových oblastí a významných objektů v popředí a rozšířili ho do víceúrovňového grafu. Yan a kol. [38] představili hierarchický model (HS) snižující efekt složitých struktur, zatímco Shi a kol. [39] rozšířili tento model o vícevrstvý přístup. Shen a Wu [40] navrhli metodu obnovení matice s nízkým hodnocením (LR), která spojuje tradiční nízkoúrovňové vlastnosti spolu s vysokoúrovňovými pro vyšší úroveň detekce významných objektů. Peng a kol. [41] představili nový model rozkladu strukturované matice (SMD) se dvěma strukturálními regularizacemi pro detekci významných objektů. Liu a kol. [42] navrhli model stromů významnosti (ST) pro její efektivní měření.

b) Modely využívající algoritmy strojového učení: Margolin a kol. [43] navrhli nový přístup k odhalování významných oblastí prostřednictvím identifikace jedinečných vzorů (PCA), zatímco Hou a kol. [44] navrhli spektrální reziduální (SR) přístup. Wang a kol. [45, 46] představili integraci kontrolovaných atributů k diskriminační integraci významných prvků pro výpočet významnosti. He a kol. [47] představili novou metodu, která používá konvoluční neuronovou síť (CNN) zaměřenou na superpixely k detekci významných oblastí. Zhao a kol. [48] využili techniky hlubokého učení k detekci významnosti, přičemž jak globální tak lokální kontext jsou brány v úvahu. Chen a kol. [49] využili metodu deep-CNN k progresivnímu učení reprezentací významnosti a Li a kol. [50] navrhli víceúrovňový model významnosti založený na plně konvoluční neuronové síti. Li a Yu [51] představili koncovou síť hlubokého kontrastu pro detekci významných objektů. Liu a Han [52] navrhli hlubokou hierarchickou síť pro detekci významnosti.

Výše uvedené metody jsou některé z těch nejmodernějších metod detekce významnosti. Toet [53] a Borji a Itti [54] předložili přehled stávajících modelů významnosti.

Vytvoření metody, která má vysokou výpočetní efektivitu a generuje vysoce kvalitní mapy významnosti však zůstává náročné.

3 Výběr metod

V kapitole 2 byla představena řada metod pro detekci významných objektů a oblastí. V příštích kapitolách budou popsány tři metody z tří různých kategorií.

Tyto tři metody jsme si vybrali, jelikož představují tři různé druhy způsobů řešení problému detekce významných oblastí v obrazech a výborně znázorňují výhody a nevýhody jednotlivých druhů.

Metoda histogramového kontrastu

Tato metoda, zastupující kategorii *zdola nahoru* bude představena v kapitole 4. Principem této metody je přiřazení hodnot významnosti jednotlivým barvám v obraze pomocí porovnání barevného kontrastu. Tento způsob je účinný při detekci významnosti v obrazech, kde významná oblast sestává z barev, které se výrazně odlišují od pozadí jak kontrastem, tak barvou.

Metoda minimální bariérové vzdálenosti

Metoda MBD, představena v kapitole 5, je zastupitelem kategorie *shora dolů*, konkrétně pak podkategorie *modelů využívajících vysokoúrovňové vlastnosti*. Tato metoda je založena na principu, že oblasti pozadí jsou obvykle spojeny s okrajem obrazu. Pro efektivní výpočet tohoto spojení je použit algoritmus rastrového skenování. Tento princip má však jednu výraznou nevýhodu - nedokáže rozpoznat významné oblasti, které se dotýkají okraje. Tato skutečnost je podrobněji zkoumána v kapitole “objekty dotýkající se okrajů” na straně 34.

Metoda spektrálního rezidua

V kapitole 6 bude popsána metoda SR, která podobně jako metoda MBD patří do kategorie *shora dolů*, ale na rozdíl od ní spadá do podkategorie *modelů využívajících algoritmy strojového učení*. Spektrální reziduum je rozdílem logaritmického spektra a zprůměrovaného logaritmického spektra obrazu, obsahuje tedy části spektra, které se vymykají průměru. Tyto části korespondují s inovacemi v obraze, což znamená, že metoda SR je výborná pro detekci oblastí, kde se rychle mění barva a kontrast.

4 Metoda histogramového kontrastu

Náš biologický vizuální systém je vysoce citlivý na kontrast. Na základě tohoto pozorování navrhli Cheng a kol. [24, 25, 26] metodu histogramového kontrastu (HC) pro definování hodnot významnosti u jednotlivých pixelů. Tato metoda definuje významnost pixelu pomocí barevného kontrastu tohoto pixelu a všech ostatních pixelů v obraze. Můžeme tedy říci, že významnost pixelu I_k v obraze I je

$$S(I_k) = \sum_{\forall I_i \in I} D(I_k, I_i) \quad (1)$$

kde $D(I_k, I_i)$ je metrika barevné vzdálenosti pixelů I_k a I_i v barevném prostoru $L^*a^*b^*$. Rovnice (1) může být rozepsána do podoby

$$S(I_k) = D(I_k, I_1) + D(I_k, I_2) + \dots + D(I_k, I_N) \quad (2)$$

kde N je počet pixelů v obraze I . Můžeme si všimnout, že pixely stejné barvy mají stejnou významnost, což nám umožní seskupit termíny se stejnou barvou c_j v rovnici (2) dohromady. Tímto způsobem dostaneme hodnotu významnosti pro jednotlivé barvy:

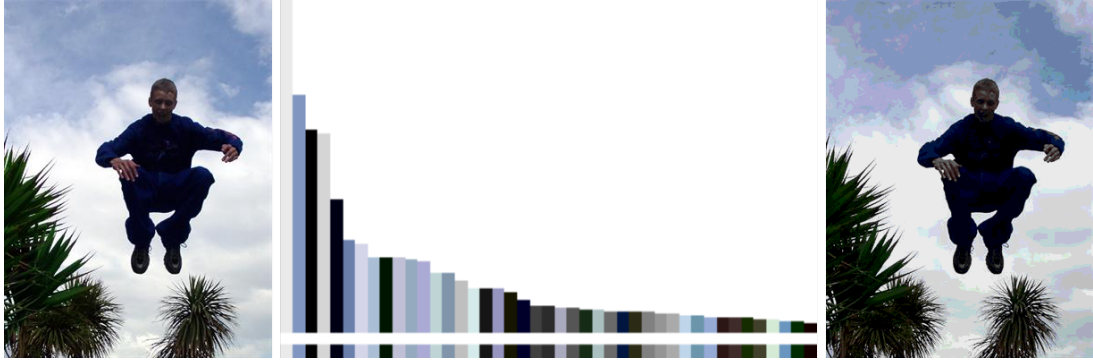
$$S(I_k) = S(c_l) = \sum_{j=1}^n f_j D(c_l, c_j) \quad (3)$$

kde c_l je barva pixelu I_k , n je počet různých barev v obraze I , f_j je pravděpodobnost výskytu barvy c_j v obraze I , a $D(c_l, c_j)$ je metrika barevné vzdálenosti barev c_l a c_j v barevném prostoru $L^*a^*b^*$.

4.1 Zrychlení založené na histogramu

Vyhodnocení hodnoty významnosti pro každý obrazový pixel pomocí rovnice (1) trvá $O(N^2)$, což je výpočetně příliš drahé i pro středně velké obrázky. Ekvivalentní reprezentace v rovnici (3) však trvá $O(N) + O(n^2)$, což znamená, že výpočetní účinnost může být zlepšena na $O(N)$, pokud $O(n^2) \leq O(N)$. Klíčem k urychlení je tedy snížení počtu barev v obraze. Barevný prostor RGB však obsahuje 256^3 možných barev, což je obvykle více než počet pixelů.

Zhai a Shah [55] snižují počet barev n používáním pouhého jasu. Tímto způsobem je $n^2 = 256^2$ (obvykle $256^2 \ll N$). Tato metoda však ignoruje odlišnost informací v celém barevném spektru. V této práci používáme celý barevný prostor namísto jasu. Abychom snížili počet barev, které je třeba vzít v úvahu, nejprve kvantifikujeme každý barevný kanál tak, aby měl pouze 12 různých hodnot, což snižuje počet barev na $12^3 = 1728$. Vzhledem k tomu, že barva v přirozeném obraze obvykle pokrývá pouze malou část celého barevného prostoru, dále snižujeme počet barev ignorováním méně často se vyskytujících barev. Výběrem častěji se vyskytujících barev a zajištěním, že tyto barvy pokrývají více než 95% obrazových pixelů, nám obvykle zůstane kolem $n = 85$ barev. Barvy zbývajících pixelů, které tvoří méně než 5% obrazu, jsou v histogramu



Obrázek 1: Pro vstupní obrázek (vlevo) vypočítáme jeho barevný histogram (uprostřed). Odpovídající barvy zásobníků histogramu jsou zobrazeny ve spodní liště. Kvantifikovaný obrázek (vpravo) používá pouze 43 barev histogramu a stále si zachovává dostatečnou vizuální kvalitu pro detekci významnosti.

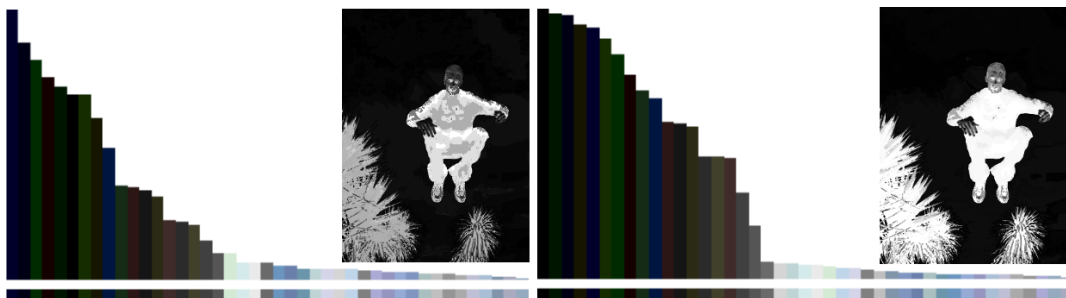
nahrazeny nejbližšími barvami. Typický příklad takové kvantifikace je znázorněn na obr. 1. Povšimněme si, že kvůli efektivitě volíme jednoduchou kvantifikaci založenou na histogramu namísto optimalizace pro specifickou barevnou paletu obrazu.

4.2 Vyhlazení barevného prostoru

Ačkoli můžeme efektivně spočítat barevný kontrast vytvořením kompaktního barevného histogramu pomocí kvantifikace barev a výběrem častějších barev, samotná kvantifikace může vytvořit artefakty. Některé podobné barvy mohou být kvantovány na různé hodnoty významnosti. Aby Cheng a kol. [26] snížili výskyt těchto anomálií, používají vyhlazovací postup k upřesnění hodnoty významnosti pro každou barvu. Tato hodnota je pro každou barvu nahrazena váženým průměrem hodnot významnosti podobných barev. Po výběru $m = n/4$ nejbližších barev můžeme upřesnit hodnotu významnosti barvy c tímto způsobem

$$S'(c) = \frac{1}{(m-1)T} \sum_{i=1}^m (T - D(c, c_i)) S(c_i) \quad (4)$$

kde $T = \sum_{i=1}^m D(c, c_i)$ je součet vzdáleností mezi barvou c a její m nejbližšími sousedy c_i , a normalizační faktor pochází z $\sum_{i=1}^m (T - D(c, c_i)) = (m-1)T$. Povšimněme si, že používáme lineárně se měnící vyhlazovací váhu $(T - D(c, c_i))$ k přiřazení větší váhy barvám blíže k c v barevném prostoru. Obr. 2 ukazuje typický účinek vyhlazení barevného prostoru s odpovídajícími histogramy seřazenými podle klesajících hodnot významnosti. Povšimněme si, že podobné zásobníky histogramu jsou po takovém vyhlazení blíže k sobě, což naznačuje, že podobným barvám bude pravděpodobně přiřazena podobná hodnota významnosti, čímž se sníží kvantifikační artefakty.



Obrázek 2: Významnost každé barvy v histogramu před (vlevo) a po (vpravo) vyhlazení barevného prostoru, spolu s odpovídajícími mapami významnosti.

4.3 Implementační detaily

Cheng a kol. [26] rovnoměrně rozdělili každý barevný kanál do 12 úrovní, aby kvantifikovali barevný prostor na 12^3 různých barev. Zatímco kvantifikování barev prováděli v barevném prostoru RGB, barevné rozdíly měřili v barevném prostoru $L^*a^*b^*$ vzhledem k jeho percepční přesnosti. Kvantifikování se neprovádí přímo v barevném prostoru $L^*a^*b^*$, protože ne všechny barvy v rozsahu $L^* \in [0, 100]$ a $a^*, b^* \in [-127, 127]$ nutně odpovídají skutečným barvám. Cheng a kol. [26] experimentálně pozorovali horší kvantifikační artefakty pomocí přímé kvantifikace barevného prostoru $L^*a^*b^*$. Nejlepší výsledky byly získány kvantifikací v RGB prostoru a při měření vzdálenosti v barevném prostoru $L^*a^*b^*$, na rozdíl od provádění kvantifikace a výpočtu vzdálenosti v jediném barevném prostoru, buď RGB nebo $L^*a^*b^*$.

5 Metoda minimální bariérové vzdálenosti

Vzhledem k rozvíjejícím se aplikacím na mobilních zařízeních a rozsáhlých datasetech by žádoucí metoda detekce významných objektů neměla produkovat pouze vysoce kvalitní mapy významnosti, ale měla by být také vysoce výpočetně efektivní. Zhang a kol. [35] ve svém článku představili algoritmus pro transformaci přibližné minimální bariérové vzdálenosti (MBD) pro detekci významných oblastí a zabývali se jak požadavky na kvalitu, tak na rychlost detekce významných objektů.

Bylo dokázáno, že teorie připojení okraje obrazu (Image Boundary Connectivity Cue), která předpokládá, že oblasti pozadí jsou obvykle spojeny s okrajem obrazu, je efektivní pro detekci významných objektů [31, 29, 36, 33]. Aby bylo možné tuto teorii využít, předchází metody založené na geodetické vzdálenosti [31, 33] nebo na difuzi [36, 32] se při extrakci superpixelů opírají o krok abstrakce oblasti. Reprezentace superpixelů pomáhá odstranit irelevantní detaily obrazů a/nebo činí tyto modely výpočtově proveditelné. Tento krok abstrakce oblasti se však také stává rychlostní překážkou pro tento typ metod.

Pro zvýšení rychlosti navrhuje Zhang a kol. [35] metodu využití připojení okraje obrazu bez abstrakce oblasti. K měření připojení pixelu k okraji obrazu používají minimální bariérovou vzdálenost (MBD) [56, 57]. Ve srovnání s široce používanou geodetickou vzdáleností je MBD mnohem odolnější vůči kolísání hodnot pixelů.

Protože přesný algoritmus pro transformaci MBD není příliš efektivní, představili Zhang a kol. [35] FastMBD, rychlý rastrový skenovací algoritmus pro transformaci MBD, který poskytuje dobrou aproximaci transformace MBD v rámci milisekund, což je o dva řády rychlejší než přesný algoritmus [57].

5.1 Transformace vzdálenosti

Transformace vzdálenosti obrazu se zaměřuje na výpočet mapy vzdálenosti s ohledem na množinu počátečních pixelů. Jako velmi mocný nástroj pro geometrickou analýzu obrazů se jedná o dlouho studované téma v počítačovém vidění [58].

Formálně uvažujeme 2-D jednokanálový digitální obraz \mathcal{I} . Cesta $\pi = \langle \pi(0), \dots, \pi(k) \rangle$ v obraze \mathcal{I} je posloupnost pixelů, kde po sobě jdoucí dvojice pixelů sousedí. Pro určení hodnoty významnosti obrazového pixelu t potřebujeme podle teorie připojení okraje obrazu najít nejkratší cestu π spojující pixel t s množinou počátečních pixelů \mathcal{S} . Když uvažujeme nákladovou funkci \mathcal{F} a množinu počátečních pixelů \mathcal{S} , problém transformace vzdálenosti vyžaduje výpočet mapy vzdálenosti \mathcal{D} , takže pro každý pixel t

$$\mathcal{D}(t) = \min_{\pi \in \Pi_{\mathcal{S},t}} \mathcal{F}(\pi) \quad (5)$$

kde $\Pi_{\mathcal{S},t}$ je množina všech cest, které spojují pixel t s pixely v množině \mathcal{S} .

Definice nákladové funkce \mathcal{F} je závislá na aplikaci. V [31, 33] je použita geodetická vzdálenost pro detekci významných objektů. Když uvažujeme jednokanálový obraz \mathcal{I} je geodetická nákladová funkce $\Sigma_{\mathcal{I}}$ definována jako

$$\Sigma_{\mathcal{I}}(\pi) = \sum_{i=1}^k |\mathcal{I}(\pi(i-1)) - \mathcal{I}(\pi(i))| \quad (6)$$

kde $\mathcal{I}(\cdot)$ označuje hodnotu pixelu. Geodetická vzdálenost je tedy součtem všech rozdílů mezi sousedícími pixely v cestě π . V [56] byla navržena nová nákladová funkce:

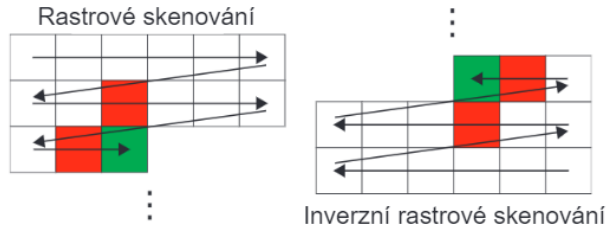
$$\beta_{\mathcal{I}}(\pi) = \max_{i=0}^k \mathcal{I}(\pi(i)) - \min_{i=0}^k \mathcal{I}(\pi(i)) \quad (7)$$

Tato vzdálenost se nazývá minimální bariérová vzdálenost (MBD) a bylo dokázáno, že při segmentaci obrazů je odolnější vůči šumu a rozmazání než geodetická vzdálenost [56, 57]. Stručně řečeno je MBD rozdílem mezi největší a nejmenší hodnotou pixelů v cestě π . Přesný algoritmus pro transformaci MBD však vyžaduje časovou složitost $O(mn \log n)$ [57], kde n je počet pixelů v obraze a m je počet odlišných hodnot pixelů. V praxi může optimalizovaná implementace pro přesnou transformaci MBD trvat asi půl sekundy pro obraz 300×200 [57].

5.2 Rychlá transformace MBD pomocí rastrového skenování

Inspirováni rychlou geodetickou transformací vzdálenosti pomocí techniky rastrového skenování [59, 60], navrhuji Zhang a kol. [35] FastMBD, přibližný iterační algoritmus pro transformaci MBD. V praxi FastMBD obvykle vydává uspokojivý výsledek v několika iteracích, a proto lze říci, že má lineární složitost vzhledem k počtu obrazových pixelů. Stejně jako všechny algoritmy rastrového skenování je také šetrný k mezipaměti, takže je v praxi vysoce účinný.

Podobně jako algoritmus rastrového skenování pro geodetickou nebo euklidovskou transformaci vzdálenosti musíme během průchodu navštívit každý pixel x v rastrovém skenovacím nebo inverzním rastrovém skenovacím pořadí. Poté bude každý sousední pixel y v odpovídající polovině sousedství x (viz obr. 3) použit k iterativní minimalizaci nákladů na cestu v x tímto



Obrázek 3: Ilustrace rastrového skenovacího průchodu a inverzního rastrového skenovacího průchodu. Zelený pixel je aktuálně navštívený pixel a jeho maskovaná sousední oblast je znázorněna červeně. [35]

způsobem

$$\mathcal{D}(x) \leftarrow \min \begin{cases} \mathcal{D}(x) \\ \beta_{\mathcal{I}}(\mathcal{P}(y) \cdot \langle y, x \rangle) \end{cases} \quad (8)$$

kde $\mathcal{P}(y)$ označuje cestu aktuálně přiřazenou k pixelu y , $\langle y, x \rangle$ označuje okraj od y do x , a $\mathcal{P}(y) \cdot \langle y, x \rangle$ je cesta pro x , která připojuje okraj $\langle y, x \rangle$ k $\mathcal{P}(y)$.

Nechť $\mathcal{P}_y(x)$ označuje $\mathcal{P}(y) \cdot \langle y, x \rangle$. Povšimněme si, že

$$\beta_{\mathcal{I}}(\mathcal{P}_y(x)) = \max\{\mathcal{U}(y), \mathcal{I}(x)\} - \min\{\mathcal{L}(y), \mathcal{I}(x)\} \quad (9)$$

kde $\mathcal{U}(y)$ a $\mathcal{L}(y)$ jsou nejvyšší a nejnižší hodnoty pixelů na $\mathcal{P}(y)$. Novou nákladovou funkci $\beta_{\mathcal{I}}(\mathcal{P}_y(x))$ lze proto efektivně vypočítat pomocí dvou pomocných map \mathcal{U} a \mathcal{L} , které sledují nejvyšší a nejnižší hodnoty na aktuální cestě pro každý pixel.

Algoritmus 1: FastMBD

vstup : obraz $\mathcal{I} = (I, V)$, množina počátečních pixelů \mathcal{S} , počet průchodů K
výstup : MBD mapa \mathcal{D}
pomocné mapy: \mathcal{U}, \mathcal{L}

nastav $\mathcal{D}(x)$ na 0 pro $\forall x \in \mathcal{S}$; jinak nastav $\mathcal{D}(x)$ na ∞ ;
nastav $\mathcal{L} \leftarrow \mathcal{I}$ a $\mathcal{U} \leftarrow \mathcal{I}$;
for $i = 1 : K$ **do**
 if $\text{mod}(i, 2) = 1$ **then**
 RasterScan($\mathcal{D}, \mathcal{U}, \mathcal{L}, \mathcal{I}$);
 else
 InvRasterScan($\mathcal{D}, \mathcal{U}, \mathcal{L}, \mathcal{I}$);

Algoritmus 2: RasterScan($\mathcal{D}, \mathcal{U}, \mathcal{L}; \mathcal{I}$)

for každé x , které je navštíveno v pořadí rastrového skenování **do**
 for každé y v maskované oblasti pro x **do**
 vypočítej $\beta_{\mathcal{I}}(\mathcal{P}_y(x))$ podle rovnice (9);
 if $\beta_{\mathcal{I}}(\mathcal{P}_y(x)) < \mathcal{D}(x)$ **then**
 $\mathcal{D}(x) \leftarrow \beta_{\mathcal{I}}(\mathcal{P}_y(x))$;
 $\mathcal{U}(x) \leftarrow \max\{\mathcal{U}(y), \mathcal{I}(x)\}$;
 $\mathcal{L}(x) \leftarrow \min\{\mathcal{L}(y), \mathcal{I}(x)\}$;

Když uvažujeme obraz \mathcal{I} a množinu počátečních pixelů \mathcal{S} , je inicializace distanční mapy \mathcal{D} a pomocných map \mathcal{U} a \mathcal{L} popsána v alg. 1. Poté se střídavě aplikují dva podprogramy, rastrový skenovací průchod a inverzní rastrový skenovací průchod, pro aktualizaci \mathcal{D} a pomocných map, dokud není dosaženo požadovaného počtu průchodů (viz alg. 1). Podprogram rastrového skenování je popsán v alg. 2. Inverzní rastrový skenovací průchod je v podstatě stejný jako alg. 2,

kromě toho, že prochází pixely v obráceném pořadí a používá různé masky sousedství, jak je znázorněno na obr. 3.

Každá iterace alg. 2 odpovídajícím způsobem aktualizuje \mathcal{U} a \mathcal{L} . V každém stavu alg. 1 je tedy $\mathcal{D}(x)$ nákladem nějaké cesty, která spojuje množinu počátečních pixelů \mathcal{S} s pixelem x . Z toho vyplývá, že $\mathcal{D}(x)$ je horní mez přesné hodnoty MBD x v kterémkoli kroku. Alg. 1 konverguje, jelikož každá hodnota v \mathcal{D} je během aktualizace nezáporná a nezvyšuje se. Konvergované řešení bude také horní hranicí přesné hodnoty MBD pro každý pixel.

5.3 Transformace MBD pro detekci významných objektů

Podobně jako v [31], abychom uplatnili teorii připojení okraje obrazu, nastavíme pixely podél okraje obrazu jako množinu počátečních pixelů \mathcal{S} a vypočítáme transformaci MBD pro každý barevný kanál pomocí FastMBD. Pak jsou MBD mapy pro všechny barevné kanály sečteny po pixelech dohromady, aby vytvořily kombinovanou MBD mapu \mathcal{B} , jejíž hodnoty pixelů jsou dále upraveny tak, že maximální hodnota je 1. V alg. 1 používáme tři průchody.

5.4 Post-processing

Zhang a kol. [35] popisují řadu efektivních operací pro zvýšení kvality výsledné mapy významnosti $\mathcal{V} = \mathcal{B}$. Tyto operace nezvyšují výpočetní zátěž, ale mohou účinně vylepšit finální výsledek.

Podobně jako v mnoha předchozích metodách [40, 23], aby Zhang a kol. [35] zohlednili zkreslení středu, které je pozorováno v mnoha datasetech detekce významných objektů [61], násobí pixely \mathcal{V} s mapou soustředění $\mathcal{C} = [c^{ij}]_{W \times H}$, která je definována jako

$$c^{ij} = 1 - \frac{\sqrt{(i - \frac{H}{2})^2 + (j - \frac{W}{2})^2}}{\sqrt{(\frac{H}{2})^2 + (\frac{W}{2})^2}} \quad (10)$$

kde c^{ij} je pixel v mapě soustředění \mathcal{C} na souřadnicích i a j , H je počet řádků a W je počet sloupců v mapě významnosti \mathcal{V} .

Nakonec Zhang a kol. [35] škálují hodnoty \mathcal{V} tak, aby maximální hodnota byla 1, a aplikují na \mathcal{V} operaci zvyšování kontrastu, která pomocí logistické funkce zvyšuje kontrast mezi oblastmi popředí a pozadí:

$$f(x) = \frac{1}{1 + e^{-b(x-0.5)}} \quad (11)$$

kde b je předdefinovaný parametr pro řízení úrovně kontrastu.



Obrázek 4: Mapa významnosti před (uprostřed) a po (vpravo) post-processing operacích

6 Metoda spektrálního rezidua

Většina modelů pro detekci významných objektů se zaměřuje na shrnutí vlastností cílových objektů. Není však pravděpodobné, že by existovaly obecné vlastnosti sdílené různými kategoriemi objektů. V [44] řeší Hou a Zhang tento problém alternativním způsobem: zkoumají vlastnosti pozadí. Při analýze logaritmického spektra obrazu získají spektrální reziduum, které potom transformují na prostorovou doménu, aby získali mapu významnosti.

6.1 Od spektrálního rezidua k mapě významnosti

Systém zaměřený na minimalizaci redundantních vizuálních informací si musí být vědom statistických podobností vstupních podnětů. Proto v různých logaritmických spektrech, kde lze pozorovat značné podobnosti tvarů, si zaslouží naši pozornost informace, které se vymykají z hladkých křivek. Hou a Zhang [44] předpokládají, že statistické singularity ve spektru mohou být zodpovědné za neobvyklé oblasti v obraze, kde se objevují významné objekty.

Pro výpočet logaritmického spektra $\mathcal{L}(f)$ používáme amplitudové spektrum $\mathcal{A}(f)$ Fourierovy transformace obrazu:

$$\mathcal{L}(f) = \log(\mathcal{A}(f)) \quad (12)$$

Logaritmické spektrum $\mathcal{L}(f)$ obsahuje statistickou charakteristiku vstupního obrazu. Pro nalezení statistických singularit v této charakteristice potřebujeme nejprve najít průměrný tvar logaritmického spektra jeho rozmazáním.

Hou a Zhang [44] používají místní průměrný filtr $h_n(f)$ pro rozmazání $\mathcal{L}(f)$, kde n se rovná 3. Změna velikosti $h_n(f)$ změní výsledek jen nepatrně, viz obr. 5. Zprůměrované spektrum $\mathcal{Z}(f)$ lze získat rozmazáním logaritmického spektra:

$$\mathcal{Z}(f) = h_n(f) * \mathcal{L}(f) \quad (13)$$

kde $h_n(f)$ je $n \times n$ matice definovaná tímto způsobem:

$$h_n(f) = \frac{1}{n^2} \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix} \quad (14)$$

Pro nalezení statistických singularit v logaritmickém spektru $\mathcal{L}(f)$ od něj stačí odečíst zprůměrované spektrum $\mathcal{Z}(f)$. To, co nám zbude, jsou části logaritmického spektra, které se vymykají průměru, přičemž Hou a Zhang [44] předpokládají, že tyto části odpovídají významným oblastem obrazu. Tyto zbylé části definujeme jako spektrální reziduum $\mathcal{R}(f)$, které lze získat pomocí

$$\mathcal{R}(f) = \mathcal{L}(f) - \mathcal{Z}(f) \quad (15)$$

Spektrální reziduum obsahuje inovace v obraze. Slouží jako komprimovaná reprezentace scény. Pomocí inverzní Fourierovy transformace můžeme v prostorové doméně sestavit výstupní mapu významnosti. Tato mapa obsahuje především netriviální část scény. Obsah spektrálního rezidua lze také interpretovat jako neočekávanou část obrazu. Proto umocníme hodnotu v každém bodě mapy významnosti na druhou, aby indikovala chybu odhadu. Pro lepší vizuální efekt vyhladili Hou a Zhang [44] mapu významnosti pomocí gaussovského filtru $g(x)$ ($\sigma = 1.7$).

Stručně řečeno, pro daný obraz $\mathcal{I}(x)$ můžeme mapu významnosti $\mathcal{S}(x)$ vypočítat tímto způsobem:

$$\mathcal{A}(f) = \left| \mathfrak{F}[\mathcal{I}(x)] \right| \quad (16)$$

$$\mathcal{P}(f) = \text{angle}(\mathfrak{F}[\mathcal{I}(x)]) \quad (17)$$

$$\mathcal{L}(f) = \log(\mathcal{A}(f)) \quad (18)$$

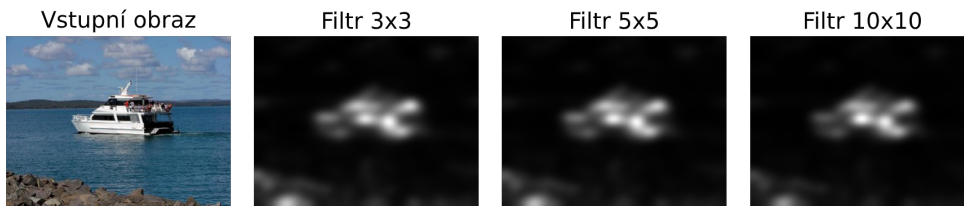
$$\mathcal{Z}(f) = h_n(f) * \mathcal{L}(f) \quad (19)$$

$$\mathcal{R}(f) = \mathcal{L}(f) - \mathcal{Z}(f) \quad (20)$$

$$\mathcal{S}(x) = g(x) * \mathfrak{F}^{-1} \left[\exp(\mathcal{R}(f) + i \cdot \mathcal{P}(f)) \right]^2 \quad (21)$$

kde \mathfrak{F} a \mathfrak{F}^{-1} označují Fourierovu transformaci a inverzní Fourierovu transformaci, $\mathcal{A}(f)$ je amplitudové spektrum, $\mathcal{P}(f)$ je fázové spektrum, $\mathcal{L}(f)$ je logaritmické spektrum, $\mathcal{R}(f)$ je spektrální reziduum, $h_n(f)$ a $g(x)$ označují průměrný a gaussovský filtr, i je imaginární jednotka, a $[\cdot]^2$ označuje operaci umocnění každé hodnoty matice na druhou.

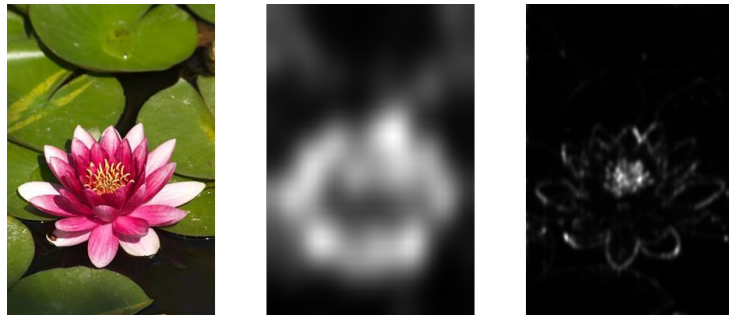
Pro daný vstupní obraz se jeho logaritmické spektrum $\mathcal{L}(f)$ počítá z obrazu škálovaného na 64×64 px. Výběr velikosti vstupu souvisí s vizuální škálou. Vztah mezi vizuální škálou a významností je diskutován v části 6.2.



Obrázek 5: Ilustrace efektu velikosti místního průměrného filtru.

6.2 Výběr vizuální škály

Vizuální systém funguje pod určitými měřítky. Například ve velkém měřítku můžeme vnímat dům jako objekt, ale v malém měřítku je velmi pravděpodobné, že budeme jako objekt vnímat spíše přední dveře domu. Výběr měřítka v našem experimentu odpovídá výběru velikosti vstupního obrazu. Když je obraz velký, vynechají se podrobné vlastnosti a vizuální vyhledávání se provádí ve velkém měřítku. V jemnějším měřítku se však větší rysy stanou méně konkurenceschopnými vůči malým, ale náhlým změnám v obraze. Změna měřítka vede k odlišnému výsledku v mapě významnosti. Tato vlastnost je znázorněna na obr. 6.

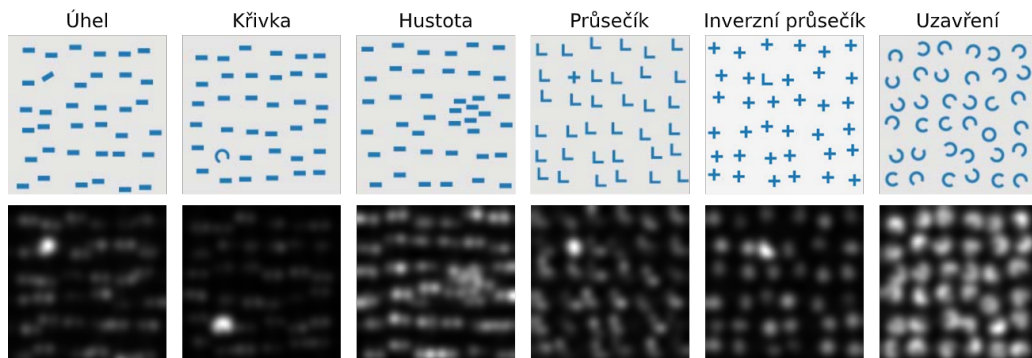


Obrázek 6: Ilustrace efektu vizuálního měřítka. Původní obraz (vlevo) je transformován na mapu významnosti po škálování na 32x32 (uprostřed) a 256x256 (vpravo). Můžeme vidět, že ve větším měřítku je jako objekt vnímám celý květ, zatímco v menším měřítku je objektem střed květu.

6.3 Reakce na psychologické vzorce

Metodu spektrálního rezidua Hou a Zhang [44] otestovali na umělých vzorcích. Tyto vzorce jsou zkoumány v řadě experimentů pozornosti [62, 63].

Všeobecně se uznává, že určité komplexní rysy přesahují možnosti podvědomého vnímání. K rozlišování singularit ve vzorcích, jako je „Uzavření“ na obr. 7, musí být použit delikátnější a časově náročnější vyhledávací proces. Z tohoto důvodu nedokáže metoda spektrálního rezidua najít kruh mezi „c“.



Obrázek 7: Reakce na psychologické vzorce [44]

7 Porovnání

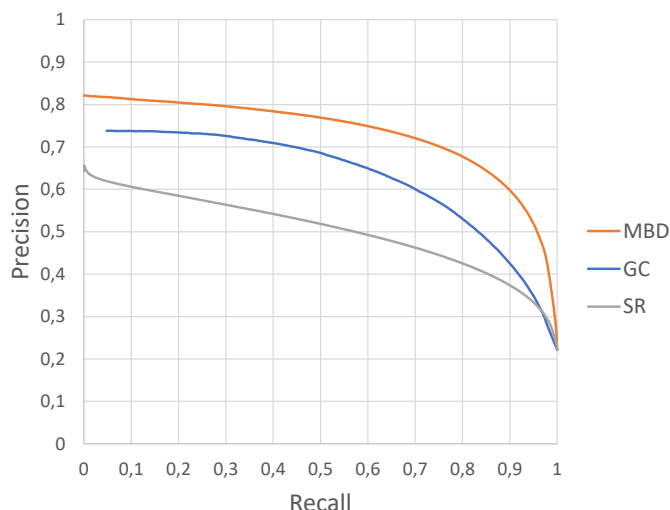
V této sekci se zaměříme na porovnání výše popsaných metod detekce významných objektů podle různých kategorií. První kategorie se bude zabývat precision-recall křivkou, druhá kategorie se zaměří na porovnání rychlosti naší implementace těchto metod, zatímco poslední kategorie bude porovnávat úspěšnost metod na různých typech obrazů.

7.1 Precision-recall křivka

K vyhodnocení celkové kvality metod pro detekci významných objektů používáme precision-recall křivky. Pro mapy významnosti vytvořené jednotlivými metodami na 10000 obrazech z datasetu MSRA10K [15] vygenerujeme množinu binárních obrazů pomocí prahování s prahovací hodnotou v rozsahu $[0, 255]$. Pro všechny takto vytvořené binární obrazy vypočítáme jejich precision a recall hodnoty při porovnání s odpovídajícími mapami významnosti vytvořenými člověkem, převzatými z datasetu MSRA10K. Pro přehledné grafické znázornění můžeme použít precision-recall křivku, ve které jsou na ose x vyneseny hodnoty recall a na ose y hodnoty precision. Na obr. 8 můžeme vidět precision-recall křivky pro jednotlivé metody, zprůměrované na celém datasetu MSRA10K podle jednotlivých prahovacích hodnot.

Hodnota precision je definována jako poměr správně detekovaných významných oblastí ke všem detekovaným oblastem, vyjadřuje tedy přesnost dané metody - jak často daná metoda detekuje pouze správné oblasti. Hodnota recall vyjadřuje poměr správně detekovaných významných oblastí vůči oblastem, které měly být detekovány, ukazuje tedy jak často daná metoda nachází všechny správné oblasti.

Podle výsledných precision-recall křivek našich implementací metod na obr. 8 můžeme vidět, že metoda minimální bariérové vzdálenosti je z těchto metod nejlepší.



Obrázek 8: Precision-recall křivky

7.2 Rychlost

Pro porovnání rychlostí našich implementací metod detekce významných oblastí jsme postupně jednotlivými metodami vypočítali mapy významnosti na 1000 obrazech o velikosti 400×300 px z datasetu MSRA10K [15]. Toto porovnání jsme prováděli na 8-vláknovém procesoru Intel Core i5-8300H @2.30 GHz.

Tabulka 1: Porovnání rychlostí jednotlivých metod - experimentální výsledky

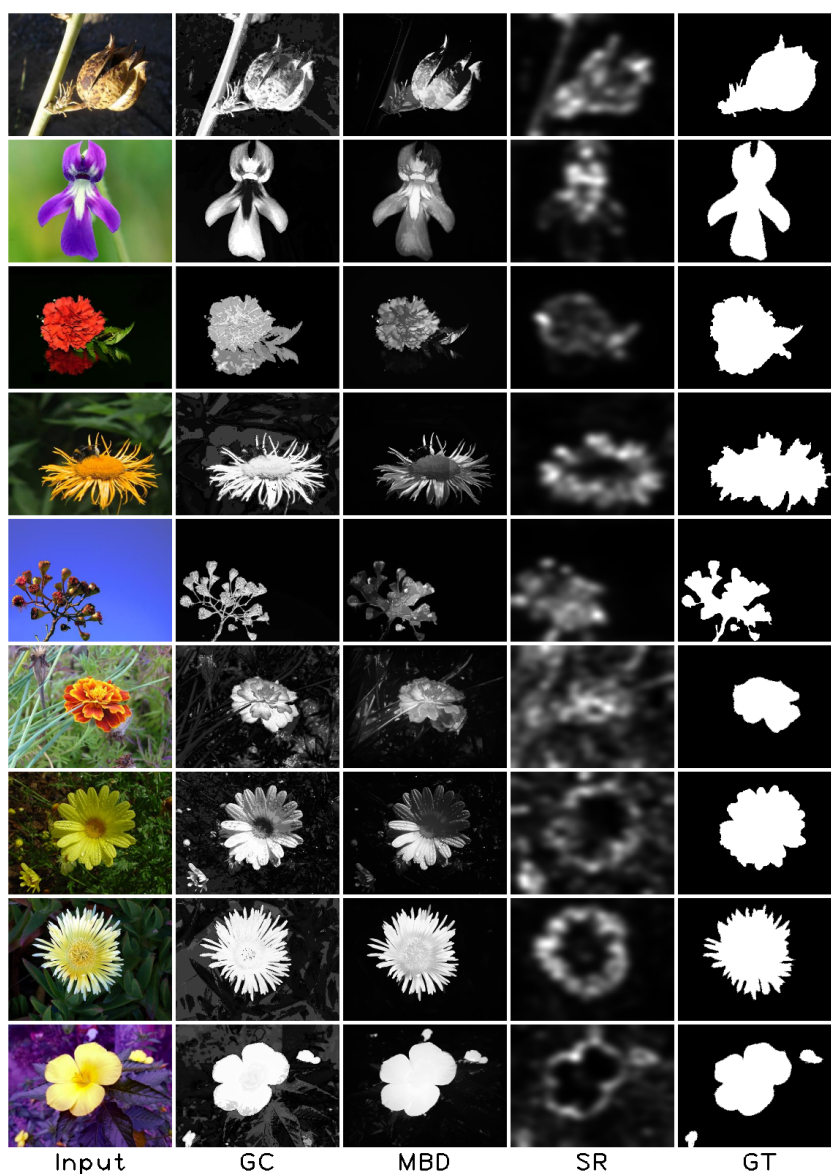
Pokus #	Metoda GC		Metoda MBD		Metoda SR	
	<i>sekundy</i>	FPS	<i>sekundy</i>	FPS	<i>sekundy</i>	FPS
1	318,98	3,1349	156,19	6,4024	7,5241	132,91
2	318,12	3,1435	158,98	6,2901	7,5075	133,20
3	319,51	3,1298	158,90	6,2933	7,5283	132,83
4	319,29	3,1319	158,08	6,3259	7,5434	132,57
5	318,14	3,1433	161,55	6,1900	7,5501	132,45
6	316,33	3,1613	158,36	6,3147	7,4547	134,14
Průměr	318,39	3,1408	158,68	6,3027	7,5180	132,50

Podle výsledků z tabulky 1 můžeme vidět že metoda spektrálního rezidua je výrazně rychlejší než zbylé dvě metody. Tento fakt můžeme vysvětlit naší neefektivní implementací těchto zbylých metod a také tím, že metoda spektrálního rezidua pracuje s obrazy zmenšenými na 64×64 px a její implementace používá z většiny knihovní metody OpenCV [64].

7.3 Různé kategorie obrazů

Tato kapitola bude porovnávat úspěšnost metod na různých kategoriích obrazů, které budou vždy rozděleny na tabulku obrazů podobného typu a na zhodnocení a porovnání map významnosti vytvořených jednotlivými metodami. Každá tabulka obrazů podobného typu je rozdělena na 5 sloupců, které postupně označují: originální obraz, mapu významnosti vytvořenou metodou histogramového kontrastu, minimální bariérové vzdálenosti, spektrálního rezidua, a nakonec sloupec s mapou významnosti vytvořenou člověkem. Originální obrazy i mapy významnosti vytvořené člověkem byly převzaty z veřejného datasetu MSRA10K.

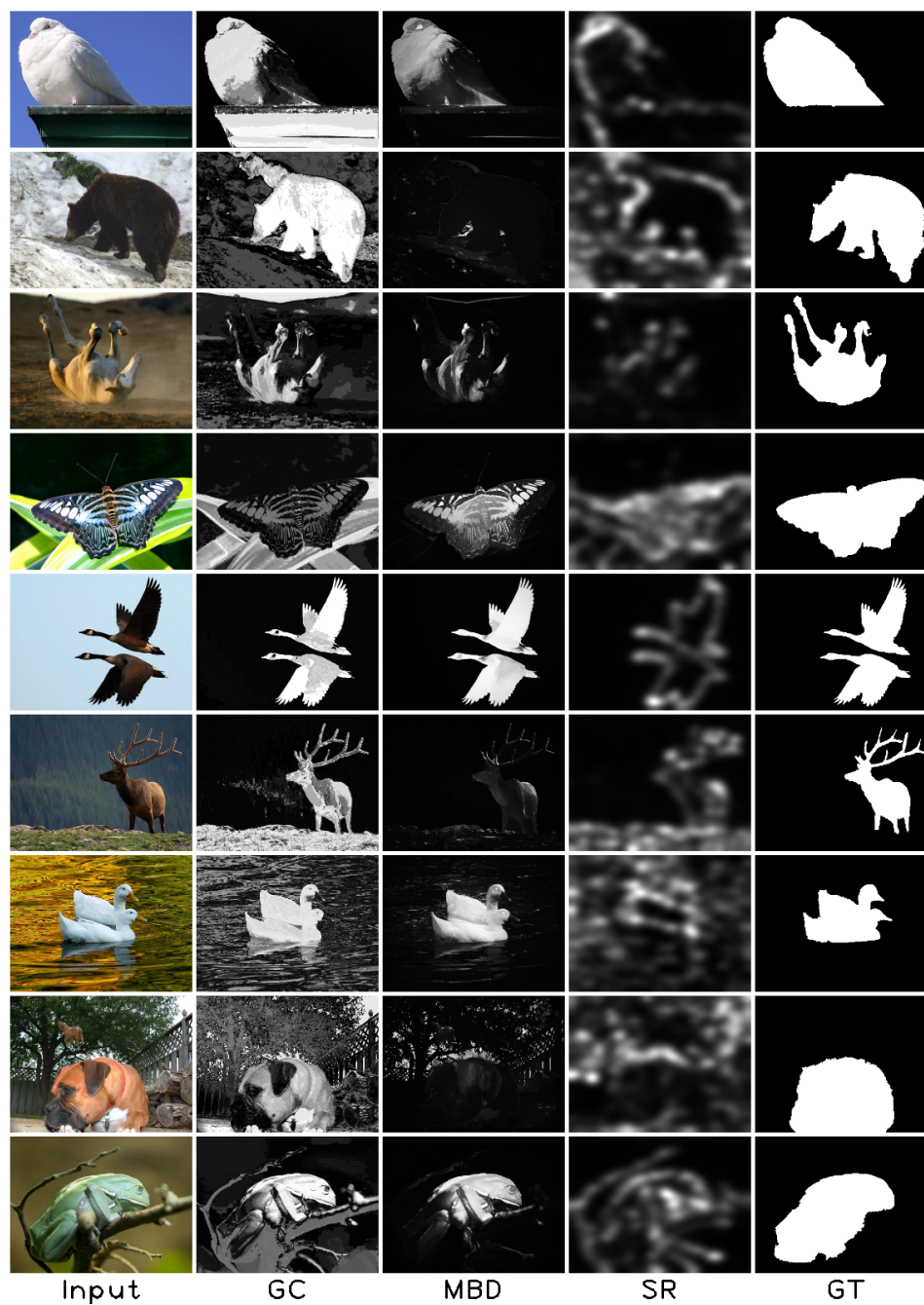
Kategorie: květiny



Obrázek 9: Porovnání metod na obrázcích květin

Květiny se svým tvarem a zbarvením obvykle podstatně odlišují od pozadí, proto si všechny tři metody radí s touto kategorií velice dobře. Metoda spektrálního rezidua má tendenci rozlišovat části obrazu kde se mění barva a kontrast, proto často označuje jako salienční oblasti pouze okraje květin. Můžeme si také povšimnout, že na prvním řádku je významný rozdíl mezi dvěma prvními metodami - metoda GC označuje jako významný stonek květiny dotýkající se okraje obrazu, zatímco metoda MBD nikoliv. Tato skutečnost je podrobněji zkoumána v kapitole “objekty dotýkající se okrajů” na straně 34.

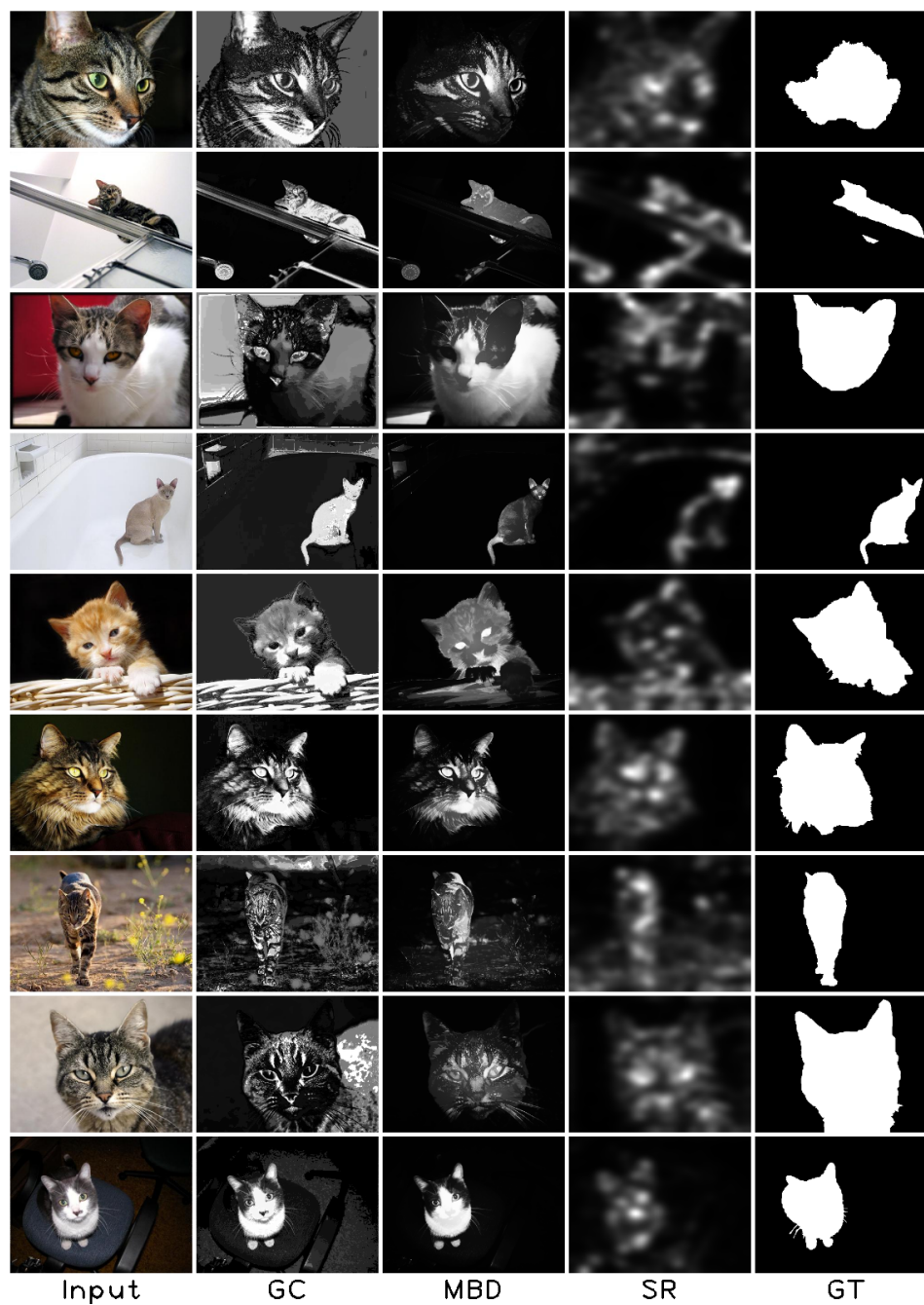
Kategorie: zvířata



Obrázek 10: Porovnání metod na obrázcích zvířat

U obrázků se zvířaty je určení významných oblastí náročné, jelikož ani jedno zvíře nebylo správně označeno všemi metodami, kromě černých kachen, které se výrazně odlišují od pozadí jak kontrastem, tak barvou.

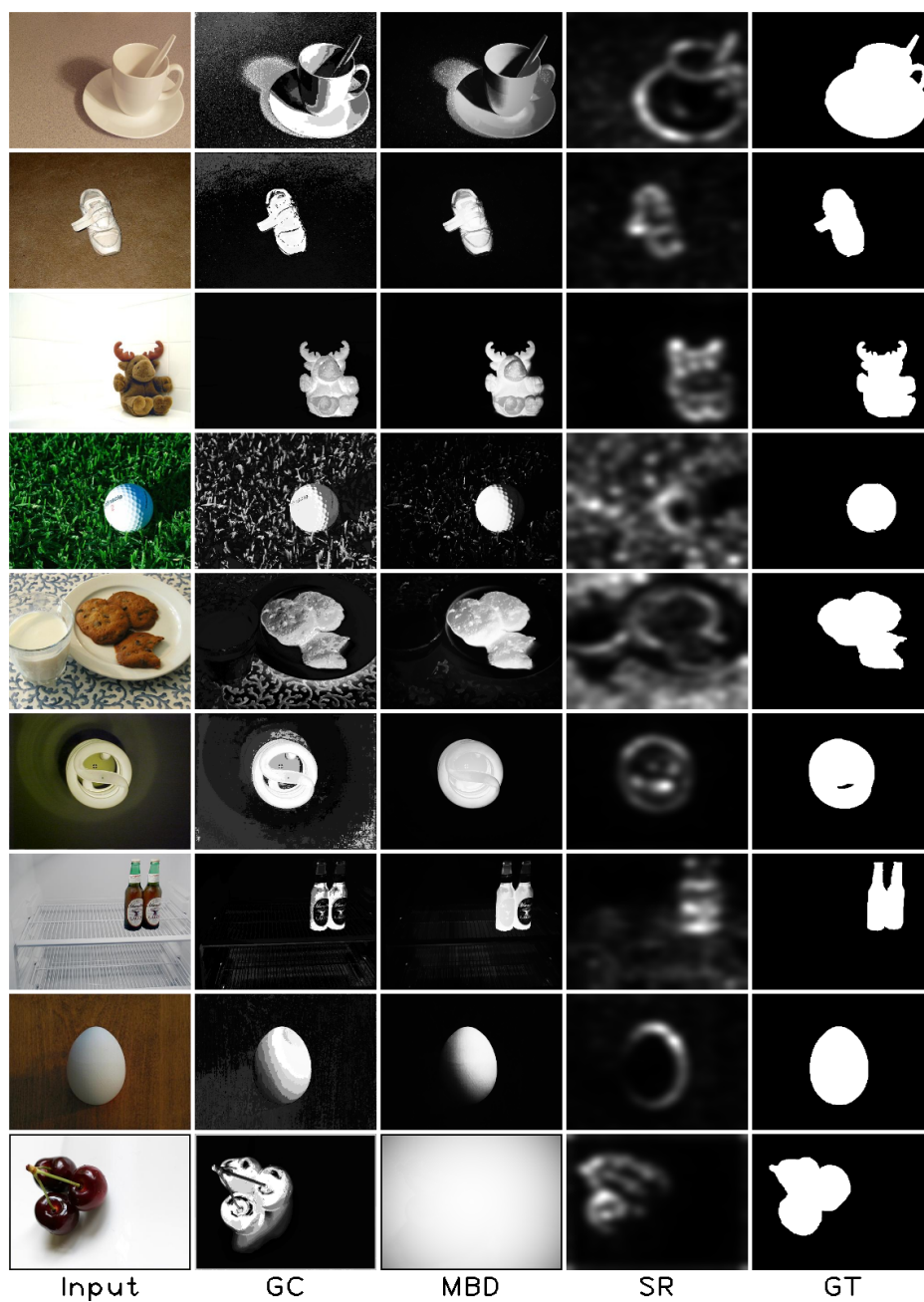
Kategorie: kočky



Obrázek 11: Porovnání metod na obrázcích koček

Kočky se prokazují být nejtěžší kategorií pro všechny tři metody. Nejenže se často podobají svému pozadí, ale také jednotlivé části koček se výrazně liší, což znemožňuje správnou identifikaci celé kočky.

Kategorie: všední objekty



Obrázek 12: Porovnání metod na obrazech všedních objektů

Určení významných oblastí u obrazů s všedními objekty je snadné pro MBD, ale i ostatní dvě metody si radí poměrně dobře. Jedinou výjimkou je poslední obraz s třešněmi, kde metoda MBD označila jako salienční oblast celý obraz kromě okrajů. Tento fakt je způsoben tenkým černým okrajem na vstupním obraze. MBD je založena na rozdílu minimálního a maximálního jasu, které byly díky tomuto okraji nalezeny hned na počátku a jasy dál od okraje už nezměnily hodnotu vzdálenosti.

Kategorie: lidé



Obrázek 13: Porovnání metod na obrazech lidí

Všechny tři metody zvládají označit jednotlivé lidské postavy na kontrastním pozadí bez větších problémů, zatímco obrazy s větším seskupením lidí jsou nesprávně označeny hlavně z důvodu velkých barevných a kontrastních rozdílů v oblečení jednotlivých lidí. Znovu si můžeme povšimnout, že metoda MBD nedokázala označit významnou oblast na obraze s člověkem dotýkajícím se okraje - viz. kapitola “objekty dotýkající se okrajů” na straně 34.

Kategorie: auta



Obrázek 14: Porovnání metod na obrazcích aut

S auty si každá metoda radí různě. Metoda MBD dokázala skvěle najít významné oblasti na posledních pěti obrazech, ale na prvních čtyřech měla problém buďto s koly a skly, nebo s karosérií auta. Metoda SR se jako obvykle zaměřila hlavně na obrysy objektů - v tomto případě na rámy a kola aut.

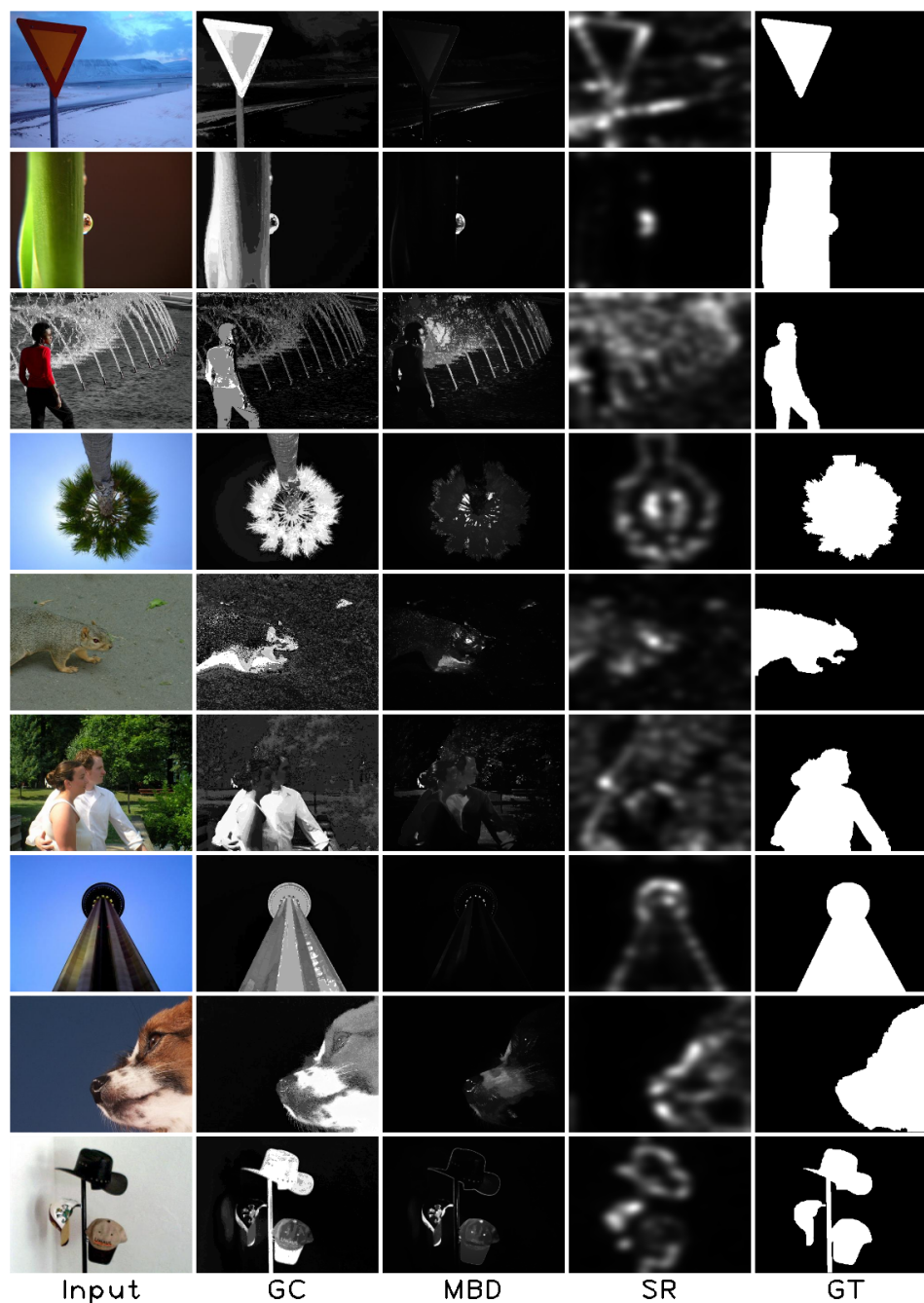
Kategorie: dopravní značky



Obrázek 15: Porovnání metod na obrázcích dopravních značek

V této kategorii si můžeme povšimnout velice důležitého rozdílu mezi metodou MBD a metodami GC a SR. Metoda MBD vnímá jako objekt celou dopravní značku, včetně nápisu na značce, který se většinou výrazně odlišuje od pozadí značky. Metoda GC dokáže rozeznat značku kromě nápisu, zatímco metoda SR vnímá jako významné nápisy i okraje značek.

Kategorie: objekty dotýkající se okrajů



Obrázek 16: Porovnání metod na obrazcích objektů dotýkajících se okrajů

Metoda MBD má jednu velice výraznou nevýhodu - předpokládá že se významná část obrazu nedotýká okraje, a že tedy jakýkoliv objekt není barevně spojen s okrajem obrazu. Tato nevýhoda způsobuje, že metoda MBD není schopná rozeznat významné objekty dotýkající se okraje jak je patrné na obr. 16.

8 Závěr

Cílem této práce bylo analyzovat a porovnat chování tří známých metod pro detekci významných oblastí a ověřit jejich úspěšnost na vybrané množině obrazů.

V první části této práce jsme představili přehled existujících metod detekce významných oblastí v obrazech, následně jsme popsali a analyzovali tři z těchto metod. První metoda se zaměřuje na využití histogramového kontrastu (HC), druhá využívá transformaci přibližné minimální bariérové vzdálenosti (MBD), a třetí aplikuje logaritmické spektrum pro získání spektrálního rezidua (SR). Na závěr jsme tyto metody otestovali na množině obrazů z veřejného datasetu MSRA10K [15] a porovnali výsledky. Metoda MBD, využívající principu připojení okraje obrazu, dosáhla při porovnání pomocí precision-recall křivek ze všech tří metod nejlepších výsledků. U všech metod jsme však našli kategorie obrazů, u kterých nejsou metody schopny správně označit významné oblasti.

Metoda MBD nedokáže kvůli využití výše zmíněného principu připojení okraje obrazu rozpoznat významné objekty dotýkající se okraje.

Metoda GC, která přiřazuje hodnoty významnosti jednotlivým barvám v obraze pomocí porovnání barevného kontrastu a četnosti výskytu, není schopna rozpoznat jako významné objekty sestávající se z více kontrastně odlišných barev, jako je tomu například u dopravních značek nebo koček.

Metoda SR využívá pro detekci významnosti logaritmické spektrum, což znamená, že detekuje oblasti, kde se rychle mění barva a kontrast. Tento princip se v praxi příliš neosvědčuje, jelikož metoda SR často označuje jako významné pouze okraje objektů, což vysvětluje, proč tato metoda dosáhla při porovnání pomocí precision-recall křivek nejhorších výsledků.

Proto usuzujeme, že tyto metody jsou vhodné pro využití v aplikacích analýzy obrazu zaměřujících se pouze na specifické typy obrazů.

Z hlediska budoucího vývoje bychom se mohli zaměřit na efektivnější implementaci jednotlivých metod, což by umožnilo porovnání více druhů metod na více datasetech. Další naskýtající se možnosti je použití lepších prostředků pro porovnávání. Precision-recall křivka je pouze jednou z možností pro zhodnocení úspěšnosti binárních klasifikačních systémů, proto bychom v budoucnu mohli využít další porovnávací techniky jako třeba ROC křivku nebo F-skóre. Mohli bychom také využít zkušeností získaných při analýze výše zmíněných metod k vytvoření nové metody pro detekci významných oblastí.

Literatura

1. HAN, Junwei; NGAN, King Ngi; LI, Mingjing; ZHANG, Hong-Jiang. Unsupervised extraction of visual attention objects in color images. *IEEE transactions on circuits and systems for video technology*. 2005, roč. 16, č. 1, s. 141–145.
2. KO, Byoung Chul; NAM, Jae-Yeal. Object-of-interest image segmentation based on human attention and semantic region clustering. *JOSA A*. 2006, roč. 23, č. 10, s. 2462–2470.
3. RUTISHAUSER, Ueli; WALTHER, Dirk; KOCH, Christof; PERONA, Pietro. Is bottom-up attention useful for object recognition? In: *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004*. 2004, sv. 2, s. II–II.
4. CHRISTOPOULOS, Charilaos; SKODRAS, Athanassios; EBRAHIMI, Touradj. The JPEG2000 still image coding system: an overview. *IEEE transactions on consumer electronics*. 2000, roč. 46, č. 4, s. 1103–1127.
5. WANG, Yu-Shuen; TAI, Chiew-Lan; SORKINE, Olga; LEE, Tong-Yee. Optimized scale-and-stretch for image resizing. In: *ACM SIGGRAPH Asia 2008 papers*. 2008, s. 1–8.
6. ZHANG, Guo-Xin; CHENG, Ming-Ming; HU, Shi-Min; MARTIN, Ralph R. A shape-preserving approach to image resizing. In: *Computer Graphics Forum*. 2009, sv. 28, s. 1897–1906. Č. 7.
7. WU, Huisi; WANG, Yu-Shuen; FENG, Kun-Chuan; WONG, Tien-Tsin; LEE, Tong-Yee; HENG, Pheng-Ann. Resizing by symmetry-summarization. *ACM Transactions on Graphics (TOG)*. 2010, roč. 29, č. 6, s. 1–10.
8. DING, Meng; TONG, Ruo-Feng. Content-aware copying and pasting in images. *The Visual Computer*. 2010, roč. 26, č. 6-8, s. 721–729.
9. TEUBER, Hans-Lukas. Physiological psychology. *Annual review of psychology*. 1955, roč. 6, č. 1, s. 267–296.
10. WOLFE, Jeremy M; HOROWITZ, Todd S. What attributes guide the deployment of visual attention and how do they do it? *Nature reviews neuroscience*. 2004, roč. 5, č. 6, s. 495–501.
11. DESIMONE, Robert; DUNCAN, John. Neural mechanisms of selective visual attention. *Annual review of neuroscience*. 1995, roč. 18, č. 1, s. 193–222.
12. MANNAN, Sabira K; KENNARD, Christopher; HUSAIN, Masud. The role of visual salience in directing eye movements in visual object agnosia. *Current biology*. 2009, roč. 19, č. 6, s. R247–R248.
13. ITTI, Laurent; KOCH, Christof; NIEBUR, Ernst. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence*. 1998, roč. 20, č. 11, s. 1254–1259.

14. ACHANTA, Radhakrishna; HEMAMI, Sheila; ESTRADA, Francisco; SUSSTRUNK, Sabine. Frequency-tuned salient region detection. In: *2009 IEEE conference on computer vision and pattern recognition*. 2009, s. 1597–1604.
15. CHENG, Ming-Ming. *MSRA10K Salient Object Database*. 2020-01. Dostupné také z: <https://mmcheng.net/msra10k/>.
16. MA, Yu-Fei; ZHANG, Hong-Jiang. Contrast-based image attention analysis by using fuzzy growing. In: *Proceedings of the eleventh ACM international conference on Multimedia*. 2003, s. 374–381.
17. LE MEUR, Olivier; LE CALLET, Patrick; BARBA, Dominique; THOREAU, Dominique. A coherent computational approach to model bottom-up visual attention. *IEEE transactions on pattern analysis and machine intelligence*. 2006, roč. 28, č. 5, s. 802–817.
18. BORJI, Ali; ITTI, Laurent. Exploiting local and global patch rarities for saliency detection. In: *2012 IEEE conference on computer vision and pattern recognition*. 2012, s. 478–485.
19. HAREL, Jonathan; KOCH, Christof; PERONA, Pietro. Graph-based visual saliency. In: *Advances in neural information processing systems*. 2007, s. 545–552.
20. FRINTROP, Simone; WERNER, Thomas; MARTIN GARCIA, German. Traditional saliency reloaded: A good old model in new shape. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, s. 82–90.
21. ACHANTA, Radhakrishna; ESTRADA, Francisco; WILS, Patricia; SÜSSTRUNK, Sabine. Salient region detection and segmentation. In: *International conference on computer vision systems*. 2008, s. 66–75.
22. LIU, Guang-Hai; YANG, Jing-Yu; LI, ZuoYong. Content-based image retrieval using computational visual attention model. *pattern recognition*. 2015, roč. 48, č. 8, s. 2554–2566.
23. GOFERMAN, Stas; ZELNIK-MANOR, Lihi; TAL, Ayellet. Context-aware saliency detection. *IEEE transactions on pattern analysis and machine intelligence*. 2011, roč. 34, č. 10, s. 1915–1926.
24. CHENG, Ming-Ming; ZHANG, Guo-Xin; MITRA, Niloy J; HUANG, Xiaolei; HU, Shi-Min. Global contrast based salient region detection. In: *CVPR 2011*. 2011, s. 409–416.
25. CHENG, Ming-Ming; WARRELL, Jonathan; LIN, Wen-Yan; ZHENG, Shuai; VINEET, Vibhav; CROOK, Nigel. Efficient salient region detection with soft image abstraction. In: *Proceedings of the IEEE International Conference on Computer vision*. 2013, s. 1529–1536.
26. CHENG, Ming-Ming; MITRA, Niloy J; HUANG, Xiaolei; TORR, Philip HS; HU, Shi-Min. Global Contrast Based Salient Region Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2015, roč. 37, č. 3, s. 569–582.

27. PERAZZI, Federico; KRÄHENBÜHL, Philipp; PRITCH, Yael; HORNUNG, Alexander. Saliency filters: Contrast based filtering for salient region detection. In: *2012 IEEE conference on computer vision and pattern recognition*. 2012, s. 733–740.
28. KIM, Jiwhan; HAN, Dongyoon; TAI, Yu-Wing; KIM, Junmo. Salient region detection via high-dimensional color transform. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, s. 883–890.
29. ZHANG, Jianming; SCLAROFF, Stan. Saliency detection: A boolean map approach. In: *Proceedings of the IEEE international conference on computer vision*. 2013, s. 153–160.
30. LI, N; YE, J; JI, Y; LING, H; YU, J. Saliency Detection on Light Field. *IEEE transactions on pattern analysis and machine intelligence*. 2017, roč. 39, č. 8, s. 1605–1616.
31. WEI, Yichen; WEN, Fang; ZHU, Wangjiang; SUN, Jian. Geodesic saliency using background priors. In: *European conference on computer vision*. 2012, s. 29–42.
32. JIANG, Bowen; ZHANG, Lihe; LU, Huchuan; YANG, Chuan; YANG, Ming-Hsuan. Saliency detection via absorbing markov chain. In: *Proceedings of the IEEE international conference on computer vision*. 2013, s. 1665–1672.
33. ZHU, Wangjiang; LIANG, Shuang; WEI, Yichen; SUN, Jian. Saliency optimization from robust background detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, s. 2814–2821.
34. JIANG, Huaizu; WANG, Jingdong; YUAN, Zejian; LIU, Tie; ZHENG, Nanning; LI, Shipeng. Automatic salient object segmentation based on context and shape prior. In: *BMVC*. 2011, sv. 6, s. 9. Č. 7.
35. ZHANG, Jianming; SCLAROFF, Stan; LIN, Zhe; SHEN, Xiaohui; PRICE, Brian; MECH, Radomir. Minimum barrier salient object detection at 80 fps. In: *Proceedings of the IEEE international conference on computer vision*. 2015, s. 1404–1412.
36. YANG, Chuan; ZHANG, Lihe; LU, Huchuan; RUAN, Xiang; YANG, Ming-Hsuan. Saliency detection via graph-based manifold ranking. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2013, s. 3166–3173.
37. ZHANG, Lihe; YANG, Chuan; LU, Huchuan; RUAN, Xiang; YANG, Ming-Hsuan. Ranking saliency. *IEEE transactions on pattern analysis and machine intelligence*. 2016, roč. 39, č. 9, s. 1892–1904.
38. YAN, Qiong; XU, Li; SHI, Jianping; JIA, Jiaya. Hierarchical saliency detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2013, s. 1155–1162.
39. SHI, Jianping; YAN, Qiong; XU, Li; JIA, Jiaya. Hierarchical image saliency detection on extended CSSD. *IEEE transactions on pattern analysis and machine intelligence*. 2015, roč. 38, č. 4, s. 717–729.

40. SHEN, Xiaohui; WU, Ying. A unified approach to salient object detection via low rank matrix recovery. In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*. 2012, s. 853–860.
41. PENG, Houwen; LI, Bing; LING, Haibin; HU, Weiming; XIONG, Weihua; MAYBANK, Stephen J. Salient object detection via structured matrix decomposition. *IEEE transactions on pattern analysis and machine intelligence*. 2016, roč. 39, č. 4, s. 818–832.
42. LIU, Zhi; ZOU, Wenbin; LE MEUR, Olivier. Saliency tree: A novel saliency detection framework. *IEEE Transactions on Image Processing*. 2014, roč. 23, č. 5, s. 1937–1952.
43. MARGOLIN, Ran; TAL, Ayellet; ZELNIK-MANOR, Lihi. What makes a patch distinct? In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2013, s. 1139–1146.
44. HOU, Xiaodi; ZHANG, Liqing. Saliency detection: A spectral residual approach. In: *2007 IEEE Conference on computer vision and pattern recognition*. 2007, s. 1–8.
45. JIANG, Huaizu; WANG, Jingdong; YUAN, Zejian; WU, Yang; ZHENG, Nanning; LI, Shipeng. Salient object detection: A discriminative regional feature integration approach. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2013, s. 2083–2090.
46. WANG, Jingdong; JIANG, Huaizu; YUAN, Zejian; CHENG, Ming-Ming; HU, Xiaowei; ZHENG, Nanning. Salient Object Detection: A Discriminative Regional Feature Integration Approach. *International Journal of Computer Vision*. 2017, roč. 123, č. 2, s. 251–268.
47. HE, Shengfeng; LAU, Rynson WH; LIU, Wenxi; HUANG, Zhe; YANG, Qingxiong. Super-cnn: A superpixelwise convolutional neural network for salient object detection. *International journal of computer vision*. 2015, roč. 115, č. 3, s. 330–344.
48. ZHAO, Rui; OUYANG, Wanli; LI, Hongsheng; WANG, Xiaogang. Saliency detection by multi-context deep learning. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015, s. 1265–1274.
49. CHEN, Tianshui; LIN, Liang; LIU, Lingbo; LUO, Xiaonan; LI, Xuelong. Disc: Deep image saliency computing via progressive representation learning. *IEEE transactions on neural networks and learning systems*. 2016, roč. 27, č. 6, s. 1135–1149.
50. LI, Xi; ZHAO, Liming; WEI, Lina; YANG, Ming-Hsuan; WU, Fei; ZHUANG, Yueting; LING, Haibin; WANG, Jingdong. Deepsaliency: Multi-task deep neural network model for salient object detection. *IEEE transactions on image processing*. 2016, roč. 25, č. 8, s. 3919–3930.
51. LI, Guanbin; YU, Yizhou. Deep contrast learning for salient object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, s. 478–487.

52. LIU, Nian; HAN, Junwei. Dhsnet: Deep hierarchical saliency network for salient object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, s. 678–686.
53. TOET, Alexander. Computational versus psychophysical bottom-up image saliency: A comparative evaluation study. *IEEE transactions on pattern analysis and machine intelligence*. 2011, roč. 33, č. 11, s. 2131–2146.
54. BORJI, Ali; ITTI, Laurent. State-of-the-art in visual attention modeling. *IEEE transactions on pattern analysis and machine intelligence*. 2012, roč. 35, č. 1, s. 185–207.
55. ZHAI, Yun; SHAH, Mubarak. Visual attention detection in video sequences using spatio-temporal cues. In: *Proceedings of the 14th ACM international conference on Multimedia*. 2006, s. 815–824.
56. STRAND, Robin; CIESIELSKI, Krzysztof Chris; MALMBERG, Filip; SAHA, Punam K. The minimum barrier distance. *Computer Vision and Image Understanding*. 2013, roč. 117, č. 4, s. 429–437.
57. CIESIELSKI, Krzysztof Chris; STRAND, Robin; MALMBERG, Filip; SAHA, Punam K. Efficient algorithm for finding the exact minimum barrier distance. *Computer Vision and Image Understanding*. 2014, roč. 123, s. 53–64.
58. ROSENFELD, Azriel; PFALTZ, John L. Distance functions on digital pictures. *Pattern recognition*. 1968, roč. 1, č. 1, s. 33–61.
59. DANIELSSON, Per-Erik. Euclidean distance mapping. *Computer Graphics and image processing*. 1980, roč. 14, č. 3, s. 227–248.
60. TOIVANEN, Pekka J. New geodesic distance transforms for gray-scale images. *Pattern Recognition Letters*. 1996, roč. 17, č. 5, s. 437–450.
61. BORJI, Ali; SIHITE, Dicky N; ITTI, Laurent. Salient Object Detection: A Benchmark. In: *European Conference on Computer Vision*. 2012, s. 414–429.
62. TREISMAN, Anne M; GELADE, Garry. A feature-integration theory of attention. *Cognitive psychology*. 1980, roč. 12, č. 1, s. 97–136.
63. WOLFE, Jeremy M; OLIVA, Aude; HOROWITZ, Todd S; BUTCHER, Serena J; BOMPAS, Aline. Segmentation of objects from backgrounds in visual search tasks. *Vision research*. 2002, roč. 42, č. 28, s. 2985–3004.
64. BRADSKI, G. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*. 2000.